Improved Multi-Objective Binary Fish School for Feature Selection

Mariana Macedo, Carmelo Bastos-Filho, Ronaldo Menezes

Computer Engineering Department, University of Pernambuco, Recife, Brazil
†BioComplex Laboratory, School of Computing, Florida Institute of Technology, Melbourne, USA
mgmm@ecomp.poli.br, carmelofilho@ieee.org, rmenezes@cs.fit.edu

Abstract
The Multi-Objective Binary Fish School Search (MOBFSS) algorithm was proposed to solve optimization problems with two or three conflicting objectives and operating on discrete binary variables. The original proposal revealed good accuracy but it also exhibited a high computational cost. Here, we present strategies to obtain an improved version of MOBFSS that reaches lower Pareto fronts for minimization problems at a better computational cost. We also deploy local search procedures as proposed in BMOPSO-CDRLS to find solutions closer to the optimal solution. The achieved results outperform the state-of-art algorithms BMOPSO-CDR and BMOPSO-CDRLS in feature selection problems for hyper-volume optimization. Hence, this paper contributes to the literature in Swarm Intelligence by introducing several algorithms that can be applied to improve feature selection in the context of classification programs.

Introduction

Binary problems have been overlooked by the Swarm Intelligence community; the majority of the techniques are designed for continuous problems. Even though binary optimization problems have the same general issues present in continuous and discrete optimization, there are also significant differences. For instance, changes in one feature in a binary decision variable may completely change the context, which means that the selection of an feature can strongly impact performance. Hence, an optimization algorithm such as Fish School Search (FSS) needs to carefully select features. Furthermore, convergence operators may lead to populations of similar individuals. Such situation may happen in a continuous approach but are not as prominent as in binary problems. Consequently, new multi-objective binary algorithms should consider the idiosyncrasies of binary problems to avoid premature stagnation.

This paper proposes variations of Multi-Objective Binary Fish School Search (MOBFSS) considering the limitations and goals on binary optimization problems which include the minimize the number of features used in classification; we work with three datasets from UCI Machine Learning Repository (Lichman 2013) and look at feature selection as a function of classification error. In a nutshell, we present strategies to obtain low computational cost avoiding an increase in the error classification.

Multi-Objective Binary Fish School

The Multi-Objective Binary Fish School (MOBFSS) was proposed by Macedo et al (Macedo et al. 2017) to optimize problems with both conflicting objectives and binary decision variables. It was inspired on the manipulation of binary variables in BFSS (Sardo 2013; Sardo et al. 2014), the flip of multiple dimensions in IBFSS (Carneiro and Bastos-Filho 2016), and the operators for multi-objective problems in MOFSS (Bastos-Filho and Guimaraes 2015). In MOBFSS, the search process is similar to that of FSS in which a fish school moves in a limited search space (i.e., an aquarium) aiming at finding better regions. Each fish is a simple agent represented by a position vector \(x_i(t)\) and a weight \(w_i(t)\) associated with its performance \(f(x_i(t))\) through the iterations \(t\). The movement of the school is described by three operators: individual, collective instinctive, and collective volatile.

In the individual operator, each fish chooses a random position \(n_i(t)\) by flipping at most \(S_{ind}(t)\) dimensions, each with probability \(\text{Flip}_{ind}\), and updates its weight as \(w_i(t+1) = w_i(t) + \Delta w_i(t)D_i(t)\), where \(\Delta w_i(t)\) and \(D_i(t)\) are determined by the criteria of dominance as follows (Bastos-Filho and Guimaraes 2015). If the new solution dominates the current solution or it has a higher Crowding Distance (CD) in the case they are indifferent, \(\Delta w_i(t)\) is set to \(a\) or \(b\), respectively, and the fish moves to the new position. Conversely, if the new solution is dominated by the current solution or it has a lower Crowding Distance (CD) in the case they are indifferent, \(\Delta w_i(t)\) is set to \(-a\) or \(-b\), respectively, and the fish remains in its current position. The factor \(D_i(t)\) quantifies the extent to which a solution \(i\) is dominated by other solutions \(j\) while accounting for the number \(S_j(t)\) of solutions dominated by \(j\), and it is calculated as

\[D_i(t) = 1 - \frac{R_i(t)}{\max_j[R_j(t)]},\]

where \(R_i(t) = \sum_{j \in N, j \prec i} S_j(t)\).

Next, in the collective instinctive operator, every fish that successfully found a better position \(S_{inst}\) of their dimensions to set their values according to the values chosen by the majority. For instance, a fish will deliberately change the value of a specific dimension from 0 to 1 if most fish that improved their fitness also changed the value of that specific
In this case, an improvement means that the flipped solution only if there is an improvement in the quality of the solution. The archive. The original solution is replaced by the flipped one in the external archive if the flipped solution provides the expansion or contraction of the swarm. If the opposite happens, the dimensions that have equal values should be flipped instead. The step volatile $S_{vol}$ constrains the maximum number of dimensions that can be flipped for each fish. Therefore, the collective volatile movement provides the expansion or contraction of the swarm. This operator is essential to the balance of exploration and exploitation during the optimization process.

The last operator is the turbulence, and its goal is to prevent the external archive from getting stuck to a local minimum. In the turbulence, we perform a random flip into a copy of a solution randomly chosen within the external archive. The original solution is replaced by the flipped one only if there is an improvement in the quality of the solution. In this case, an improvement means that the flipped solution dominates the original solution of the external archive. The operator turb restricts how many features can be flipped.

### Novel Versions for the MOBFSS

We have introduced several versions of MOBFSS (henceforth called M-1) as below. Note that the M-1 version includes all the operators except for local search (see Table 1).

- **MOBFSS-2 (M-2)**: M-2 was created to minimize the number of evaluations per iteration. The differences are: there is no evaluation of an individual movement, individuals always move; in the instinctive motion, the successful fish are those who improved their solution in the previous iteration. The hypothesis is that these simplifications can be done without impacting accuracy.

- **MOBFSS-1-LS (M-1-LS) and MOBFSS-2-LS (M-2-LS)**: These are the versions of M-1 and M-2, respectively; they are endowed with local searches instead of turbulence. Given that local search strongly impacted in the improvement of BMOPSO-CDR (de Souza, Prudêncio, and Barros 2014), the hypothesis here is that that MOBFSS will also benefit.

- **MOBFSS-1-WO-T (M-1-WO-T)**: M-1-WO-T removes the turbulence from M-1. The hypothesis is that any type of perturbation is fundamental to achieve good results.

- **MOBFSS-3 (M-3)**: Similar to M-2 but with the individual operators excluded. As the individual operators mainly aim at providing diversity, the hypothesis is that the volatile movement is sufficient to provide diversity in binary problems.

- **MOBFSS-3-LS (M-3-LS)**: Based on M-3 but with local search being applied instead of the turbulence operator. The hypothesis is that this operator will optimize the results.

Table 1 shows the operators utilized in each variation of the MOBFSS introduced in this paper. These new versions were created aiming at improving the performance of the original version without the need to apply several operators and hence minimizing the complexity of the algorithm. Note that when an operator is removed, the requirements and initializations of some parameters for the algorithm become unnecessary.

### Experiments & Results

We performed the experiments using 2 computers: a MacBook Pro with a 3 GHz Intel Core i7 and 16 GB 1600 MHz DDR3, running macOS Sierra version 10.12.6; and a PC Intel Xeon 3.1 GHz, 16 GB RAM, running Ubuntu 15.10 64 bit. The programming language was Java but with the inclusion of the Jmetal library (Durillo and Nebro 2011); this library provides several techniques and metrics well established in the scientific community.

We chose the Support Vector Machine (SVM) as the classifier in our experiments because of its efficiency and robustness to several datasets and fields, and its results are good enough especially when combined with other preprocessing approaches (Hearst et al. 1998). Recall that we are comparing the MOBFSS versions according to their ability for doing feature selection, meaning that the MOBFSS versions select the features and use the SVM as a classifier using the features (i.e., the SVM is the de facto fitness function in our approach). We selected 3 datasets from UCI repository (Lichman 2013). They are Wine, Ionosphere, and Sonar with 13, 34, and 60 features, respectively. The number of features makes them more or less complex; the higher the number of features the more complex is the problem. Each version and dataset requires a different feature configuration. However, as those techniques aim to solve a large number of problems, we performed a parametric analysis for each one of the three datasets, and we chose the best solution overall.

We executed 30 simulations for each experiment in each dataset. The stop criterion and the initialized values were adopted as proposed by Souza et al. (de Souza et al. 2011), where the maximum number of evaluations of the fitness function is 200,000. The variations of the MOBFSS we propose here are compared to the BMOPSO-CDR and BMOPSO-CDRLS.

Tables 2 and 3 presents the performance of each variation as well as the two aforementioned PSO versions. The tables include the mean execution time plus the standard deviation, maximum, and minimum values for three datasets used.

The Wine dataset is the easiest experiment in this paper. Because of its simplicity, the utilized configuration rarely
impacts the results. Several settings and algorithms achieve the same best solutions, and 100 iterations are enough to reach them. In Table 2, M-3-LS shows the shortest execution time in seconds. Given the variations reaches similar solutions (see Figure 1), it is recommended to utilize the fastest version, M-3-LS. Using less than six features, the error classification for Wine remains less than 0.011%, while one of the features is of particular importance for providing high accuracy.

The Ionosphere classification is a harder problem than the classification in Wine, and each algorithm presented different results. The M-3 and M-3-LS variations had the fastest execution time as displayed in Table 2. In Table 3 we observe that the maximum hypervolume (HV) was obtained by M-3-LS which is also the second fastest algorithm. In addition, BMOPSO-CDRLS, M-1-LS, M-1 and M-3-LS shows similar mean values which indicates that the usage of local search is positive, but M-1 was not strongly impacted. Although the turbulence and local search were not determinant in this dataset, their usage seems positive for the algorithms. Spacing (S) can be evaluated to identify the diversity of the Pareto Fronts. In Table 3, S is optimized by M-3, but its performance is not one of the best. Looking at the HV and S, it is possible to observe that the M-3-LS is the best option because the diversity is easier in suboptimal locations. Table 3 shows that the maximization of the Maximum Spread (MS) is accomplished by M-3-LS revealing a considerable extension of the Pareto front, and longer Pareto fronts are easier to achieve in suboptimal locations.

The statistical test for Ionosphere dataset is exhibited in Table 4. In this table, ▲ means that the row is better than the column, ▼ represents that the row is worse than the column, and – means that the results are similar. The statistical tests were done using the Wilcoxon test with a 95% confidence interval.

3-LS statistically beats the other algorithms, and it is one of the fastest algorithms. However, M-1-LS could reach the best value of HV (the best solution for the problem). The algorithms displayed in Table 4 are better than the algorithms BMOPSO-CDR, M-2, M-2-LS and M-3. Figure 1 presents the best solution found by the 30 trials, and in both the local search was positive in all the cases. Moreover, when comparing the BMOFSS versions without local search to the BMOPSO-CDRLS, the BMOFSS versions were better.
diversity was minimized by M-2, but, as mention before, it is easier to achieve better diversity with suboptimal solutions. Thus, the algorithms which achieve better solutions provide worse results for S. M-3-LS presented the best values for the extension found on Pareto fronts (MS).

Table 4: Comparison of the best algorithms for the performance of Ionosphere dataset.

<table>
<thead>
<tr>
<th>BP-CDRLS</th>
<th>MF-1</th>
<th>MF-1-WO-T</th>
<th>MF-1-LS</th>
<th>MF-3-LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP-CDRLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF-1-WO-T</td>
<td>▼</td>
<td>▲</td>
<td>▼</td>
<td>▼</td>
</tr>
<tr>
<td>MF-1-LS</td>
<td></td>
<td>▲</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF-3-LS</td>
<td>▲</td>
<td>▲</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Comparison of the best algorithms for the performance of Sonar dataset.

<table>
<thead>
<tr>
<th>BP-CDRLS</th>
<th>MF-1</th>
<th>MF-1-LS</th>
<th>MF-3-LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP-CDRLS</td>
<td>▲</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF-1</td>
<td>▼</td>
<td>▲</td>
<td></td>
</tr>
<tr>
<td>MF-1-LS</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
</tr>
<tr>
<td>MF-3-LS</td>
<td>▲</td>
<td>▲</td>
<td>▼</td>
</tr>
</tbody>
</table>

Table 5 shows the pairwise comparison of the versions using the same Wilcoxon test used before. M-1-LS is statistically better than the others, and M-3-LS also beats the others except for M-1-LS. In addition, BMOPSO-CDRLS is better than BMOPSO-CDR, M-2, M-2-LS and M-3. Figure 1 depicts the best solution for the 30 trials where clearly the best solution was achieved by M-3-LS. It is also important to note that the usage of 10-20% of the dimensions yields a 10-17% in error classification.

Conclusion

We proposed new versions for MOBFSS aiming to minimize the computational cost while maintaining performance. A new lighter version called MOBFSS-3-LS presents a good accuracy with reduced computational cost, but it did not beat MOBFSS-1-LS regarding hypervolume in all databases. The algorithms also aimed to avoid local minima, and the local search showed significant efficiency in that regard. We contributed to Swarm Intelligence by suggesting new binary multi-objective algorithm variations in particular to the problem of feature selection. We suggest that our variations need to be analyzed considering other configurations, and tested in real-world problems.

References


Lichman, M. 2013. UCI machine learning repository.


Acknowledgments

Mariana Macedo thanks to University of Pernambuco (PFA-UPE) and CAPES (Coordination for the Improvement of Higher-Education Personnel) for the partial financial support provided. Ronaldo Menezes was partially supported by the US Army Research Office under Agreement Number W911NF-17-1-0127.